The Evaluation of Adaptable Multimodal System Outputs (ADDENDUM)

Erin Marie Panttaja
Media Lab Europe
Dublin, Ireland
{erin,reitter}@mle.media.mit.edu

David Reitter
Media Lab Europe
Dublin, Ireland

Fred Cummins
University College Dublin
Dublin, Ireland
fred.cummins@ucd.ie

Abstract

This is an addendum to the original.

1 Method

    20 users
    8 interactions
    Half distracted (per user)
    Half of the interactions good (overall)
    Half of the interactions had errors (overall)

Each interaction was a collection of system and user turns that the subject would watch and listen to on the computer. At the end of the interaction, they would be shown an email and asked whether it was the mail created in the interaction, whether they felt it was a reliable system, and whether they thought it was an efficient system (on a scale of 1-7). Distraction was provided by having the subjects play a simple flash game (on another machine).
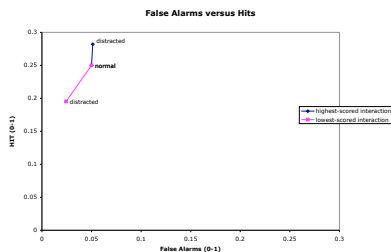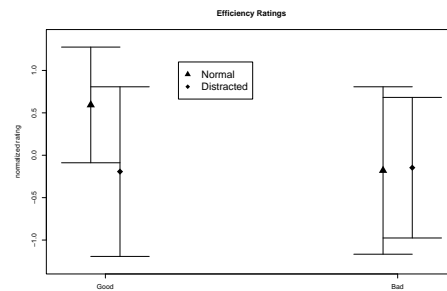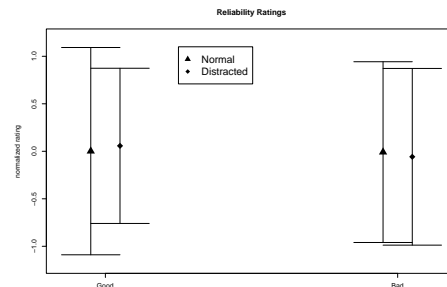

		hit	fa	normalized hit	normalized fa	
good(79)	normal	40	10	2	0.25	0.05
	distracted	39	11	2	0.282051282	0.051282051
bad (81)	normal	40	10	2	0.25	0.05
	distracted	41	8	1	0.195121951	0.024390244

summary(aov(reliab good*distr,dat))

	Df	Sum Sq	Mean Sq	F value	Pr(¿F)	
good	1	0.118	0.118	0.1302	0.7189	
distr	1	0.001	0.001	0.0008	0.9770	
good:distr	1	0.083	0.083	0.0914	0.7629	
Residuals	116	104.799	0.903			

summary(aov(effic good*distr,dat))

	Df	Sum Sq	Mean Sq	F value	Pr(¿F)	
good	1	3.676	3.676	4.6147	0.03364	*
distr	1	4.215	4.215	5.2925	0.02309	*
good:distr	1	5.344	5.344	6.7097	0.01074	*
Residuals	124	98.765	0.796			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

References

E. M. Panttaja, D. Reitter, and F. Cummins. 2004. The evaluation of adaptable multimodal system outputs. In *Proceedings of the Workshop on Robust and Adaptive Information Processing for Mobile Speech Interfaces, COLING*, Geneva, Switzerland.

2 Conclusions

In distracted situations, highly-scored interactions have a higher hit-rate than do poorly-scored interactions.

In normal situations, users perceive the more highly-scored interactions as being more efficient.

More testing will be required to further validate these results, but they are encouraging about the aptness of this evaluation methodology.